(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENT- UND MARKENAMT

(12) Übersetzung der europäischen Patentschrift

(97) EP 0 764 937 B1

(10) DE 696 13 646 T 2

(51) Int. Cl.[7]: G 10 L 11/02
G 10 L 15/20

(21) Deutsches Aktenzeichen: 696 13 646.5
(96) Europäisches Aktenzeichen: 96 115 241.0
(96) Europäischer Anmeldetag: 23. 9. 1996
(97) Erstveröffentlichung durch das EPA: 26. 3. 1997
(97) Veröffentlichungstag der Patenterteilung beim EPA: 4. 7. 2001
(47) Veröffentlichungstag im Patentblatt: 16. 5. 2002

(30) Unionspriorität:
24641895 25. 09. 1995 JP

(73) Patentinhaber:
Nippon Telegraph and Telephone Corp., Tokio/Tokyo, JP

(74) Vertreter:
Hoffmann, E., Dipl.-Ing., Pat.-Anw., 82166 Gräfelfing

(84) Benannte Vertragstaaten:
DE, FR, GB

(72) Erfinder:
Mizuno, Osamu, Sugita, Kanagawa, JP; Takahashi, Satoshi, Yokosuka-shi, Kanagawa, JP; Sagayama, Shigeki, Hoya-shi, Tokyo, JP

(54) Verfahren zur Sprachdetektion bei starken Umgebungsgeräuschen

DE 696 13 646 T 2

EP 0 764 937

Die vorliegende Erfindung betrifft ein Sprach-Endpunkt-Erfassungsverfahren und insbesondere ein Verfahren zum Erfassen einer Sprachperiode in einem Sprache enthaltenden Signal bei starken Umgebungsgeräuschen.

Spracherkennungstechnologie ist heutzutage weit verbreitet. Um Sprache zu erkennen, ist es notwendig, eine zu erkennende Sprechperiode im Eingangssignal zu erfassen. Es wird eine Beschreibung einer herkömmlichen Technik zum Erfassen der Sprechperiode auf Grundlage der Amplitude, d.h. der Leistung, der Sprache gegeben. Die hier erwähnte Leistung ist die Quadratsumme des Eingangssignals pro Zeiteinheit. Sprache enthält üblicherweise eine Tonhöhenfrequenzkomponente, deren Leistung in einer Vokalperiode besonders hoch ist. Unter der Annahme, daß ein Rahmen im Eingangssignal, in dem die Leistung des Eingangssignals einen bestimmten Schwellwert überschreitet, ein Rahmen eines Vokals ist, erfaßt das herkömmliche Schema als Sprachperiode den Vokalrahmen zusammen mit mehreren vorhergehenden und nachfolgenden Rahmen. Bei diesem Verfahren ergibt sich jedoch ein Problem, daß Signale mit hoher Leistung, die ungefähr genau so lang wie ein Wort andauern, alle irrtümlich als Sprache erfaßt werden. Das heißt, Geräusche hoher Leistung wie etwa das Geräusch einer Telefonklingel und einer zuschlagenden Tür werden als Sprache erfaßt. Ein anderes Problem dieses Verfahrens ist, daß es um so schwieriger wird, die Leistungsperiode der Sprache zu erfassen, je stärker die Leistung des Hintergrundgeräusches zunimmt. Zum Beispiel bei der Sprachsteuerung eines Instrumentes in einem Fahrzeug besteht die Möglichkeit, daß das Instrument aufgrund eines Erkennungsfehlers unkontrollierbar wird oder versagt.

Ein anderes herkömmliches Verfahren ist, die Sprachperiode auf der Basis einer Tonhöhenfrequenz zu erfassen, die die Grundfrequenz der Sprache ist. Dieses Verfahren nutzt die Tatsache, daß die Tonhöhenfrequenz eines stationären Teiles eines Vokals in den Bereich von etwa 50 bis 500 Hz fällt. Die Tonhöhenfrequenz des Eingangssignals wird untersucht, und dann wird der Rahmen, in dem die Tonhöhenfrequenz in dem oben erwähnten Frequenzbereich bleibt, als Rahmen eines Vokals angenommen, und der Rahmen sowie mehrere vorangehende und nachfolgende Rahmen werden als eine Sprachperiode erfaßt. Bei diesem Verfahren wird jedoch ein Signal mit Tonhöhenfrequenz in dem Frequenzbereich irrtümlich als Sprache erfaßt, auch wenn es ein Geräusch ist. In einer Umgebung, wo Musik mit einer im allgemeinen starken Tonkomponente einen Hintergrund bildet, ist es sehr wahrscheinlich, daß die Sprachperiode aufgrund der Tonkomponente des Musikgeräusches fehlerhaft erfaßt wird. Da außerdem das Tonhöhenfrequenz-Erfassungsverfahren die Tatsache ausnutzt, daß die Schwingungsform menschlicher Sprache bei jeder Tonhöhe eine hohe Korrelation annimmt, macht es die Überlagerung von Geräuschen über Sprache unmöglich, einen hohen Korrelationswert zu erreichen und damit die korrekte Tonhöhenfrequenz zu erfassen, was zu einem Versagen der Spracherfassung führt.

In der japanischen Patentoffenlegungsschrift Nr. 200300/85 wird ein Verfahren vorgeschlagen, das darauf abzielt, die Genauigkeit des Erfassens von Start- und Endpunkten der Sprachperiode zu verbessern. Dieses Verfahren definiert als Start- und Endpunkte der Sprachperiode diejenigen Zeitpunkte, an denen das Signalspektrum starke Veränderungen erfährt, in der Umgebung der Start- und Endpunkte einer Periode, in der die Leistung des Eingangs-Sprachsignals einen Schwellwert übersteigt. Da dieses Verfahren auf der Erfassung des Leistungspegels des Eingangssignals beruht, das den Schwellwert überschreitet, gibt es eine sehr starke Möglichkeit eines Erfassungsfehlers, der auftritt, wenn der Sprachsignalpegel niedrig oder der Geräuschpegel hoch ist.

Bei dem oben beschriebenen herkömmlichen Verfahren zum Erfassen der Sprachperiode basierend auf der Leistung der Sprache kann bei hoher Leistung des Hintergrundgeräusches dieses nicht von der Leistung der Sprache unterschieden werden, und das Geräusch wird irrtümlich als Sprache erfaßt. Andererseits gibt es bei dem Sprachperioden-Erfassungsverfahren, das auf der Tonhöhenfrequenz basiert, wenn Geräusch der Sprache überlagert wird, einen Fall, wo eine stabile Tonhöhenfrequenz nicht erhalten und deshalb Sprache nicht erfaßt werden kann. Außerdem ist in dem US-Patent Nr. 5 365 592 ein Verfahren offenbart, in dem eine Cepstrum-Tonhöhe durch eine FFT-Analyse des Eingangssignals erhalten und basierend auf der Cepstrum-Tonhöhe an jedem Zeitpunkt bestimmt wird, ob das Eingangssignal Sprache ist oder nicht. Auch dieses Verfahren ist anfällig gegen Entscheidungsfehler aufgrund von Geräuschen.

Außerdem offenbart das Dokument „Instantaneous Spectral Estimation of Nonstationary Signals" von Takizawa et al., ICASSP-94, Band IV, Seiten 329 bis 332, die Verwendung einer spektralen Frequenzänderung eines Signals für die momentane Spektralabschätzung.

Aufgabe der vorliegenden Erfindung ist daher, ein Signalverarbeitungsverfahren anzugeben, das stabile Erfassung der Sprachperiode aus dem Eingangssignal auch in einer Umgebung mit starkem Geräusch durch Ausnutzung der Informationscharakteristik von Sprache ermöglicht.

Gemäß der vorliegenden Erfindung umfaßt das Signalverarbeitungsverfahren zum Erfassen der Sprachperiode im Eingangssignal folgende Schritte:

(a) Erhalten eines spektralen Merkmalparameters durch Analysieren des Spektrums des Eingangssignals für jedes vorgegebene Analysefenster;

(b) Berechnen des Ausmaßes der Änderung des spektralen Merkmalparameters des Eingangssignals pro Zeiteinheit;

(c) Berechnen der Änderungsfrequenz des Ausmaßes des spektralen Merkmalparameters über eine vorgegebene Analyserahmenperiode, die länger als die Zeiteinheit ist; und

(d) Überprüfen, ob die Änderungsfrequenz in einen vorgegebenen Frequenzbereich fällt, und wenn ja, Entscheiden, daß das Eingangssignal des Analyserahmens ein Sprachsignal ist.

Bei dem obigen Signalverarbeitungsverfahren umfaßt der Schritt des Berechnens des Ausmaßes der Änderung des spektralen Merkmalparameters einen Schritt des Erhaltens einer Zeitfolge von Merkmalvektoren, die die Spektren des Eingangssignals an jeweiligen Zeitpunkten darstellen, und einen Schritt des Berechnens der dynamischen Messwerte durch die Verwendung der Merkmal-

vektoren an einer Mehrzahl von Zeitpunkten und des Berechnens der Änderung im Spektrum aus der Norm der dynamischen Messwerte.

Bei dem obigen Signalverarbeitungsverfahren ist der Frequenzberechnungsschritt ein Schritt des Zählens der Anzahl von Peaks der spektralen Veränderung, die einen vorgegebenen Schwellwert überschreiten und des Lieferns des resultierenden Zählergebnisses als Frequenz.

Alternativ umfaßt der Frequenzberechnungsschritt einen Schritt des Berechnens der Gesamtsumme von Änderungen im Spektrum des Eingangssignals über die Analyserahmenperiode, die länger als die Zeiteinheit ist, und der Entscheidungsschritt entscheidet, daß das Eingangssignal der Analyserahmenperiode ein Sprachsignal ist, wenn der Wert der Gesamtsumme innerhalb eines vorgegebenen Wertebereiches liegt.

Das obige Signalverarbeitungsverfahren umfaßt ferner einen Schritt des vektoriellen Quantisierens des Eingangssignals für jedes Analysefenster durch Bezugnahme auf ein Vektorcodebuch, das aus repräsentativen Vektoren für spektrale Merkmalparameter von Sprache aufgebaut ist, die aus Sprachdaten gewonnnen sind, und des Berechnens einer Quantisierungsverzerrung. Wenn die Quantisierungsverzerrung kleiner als ein vorgegebener Wert ist und die Frequenz der Änderung innerhalb des vorgegebenen Frequenzbereiches liegt, wird im Entscheidungsschritt (d) entschieden, daß das Eingangssignal im Analysefenster die Sprachperiode darstellt.

Das obige Signalverarbeitungsverfahren umfaßt ferner einen Schritt des Erhaltens der Tonhöhenfrequenz, des Amplitudenwertes oder des Korrelationswertes des Eingangssignals für jedes Analysefenster und des Entscheidens, ob das Eingangssignal ein Vokal ist. Wenn der Vokal erfaßt wird und die Frequenz der Änderung im vorgegebenen Frequenzbereich ist, wird im Entscheidungsschritt (d) entschieden, daß das Eingangssignal im Analysefenster ein Sprachsignal ist. Alternativ wird im Entscheidungsschritt (d) die Zahl von Nulldurchgängen des Eingangssignals gezählt, und basierend auf dem Zählwert wird entschieden, ob das Eingangssignal ein Konsonant ist, und wird die Sprachperiode auf der Grundlage des Entscheidungsergebnisses und der Änderungsfrequenz entschieden.

Da gemäß der vorliegenden Erfindung die Aufmerksamkeit auf die Frequenz einer spektralen Änderungscharakteristik eines Sprachtones konzentriert ist, kann sogar ein Geräusch von hoher Leistung von Sprache unterschieden werden, wenn es keine spektrale Veränderung mit der gleichen Frequenz wie die Sprache erfährt. Folglich ist es möglich, festzustellen, ob unbekannte Eingabesignale von hoher Leistung wie etwa ein stetiges Geräusch und ein sanfter Klang von Musik, Sprache sind. Auch wenn dem Sprachsignal Geräusch überlagert ist, kann Sprache mit hoher Genauigkeit erfaßt werden, weil die spektrale Änderung des Eingangssignals genau und stabil erfaßt werden kann. Außerdem können eine leise singende Stimme und andere Signale mit relativ niedriger Frequenz der spektralen Änderung beseitigt oder unterdrückt werden.

Das obige Verfahren basiert lediglich auf der Frequenz der spektralen Änderung des Eingangssignals, die Sprachperiode kann aber mit höherer Genauigkeit erfaßt werden durch Kombinieren der Frequenz der spektralen Veränderung mit ein oder mehr Informationsstücken über den

spektralen Merkmalparameter, die Tonhöhenfrequenz, den Amplitudenwert und die Zahl der Nulldurchgänge des Eingangssignals, die dessen spektrale Umhüllende zu jedem Zeitpunkt darstellen.

Fig. 1 ist ein Graph, der die Frequenz der spektralen Änderung eines Sprachsignals zeigt, auf der die vorliegende Erfindung basiert;

Fig. 2 ist ein Diagramm zur Erläuterung einer Ausgestaltung der vorliegenden Erfindung;

Fig. 3 ist ein Zeitdiagramm einer Spektralanalyse eines Signals;

Fig. 4 ist ein Diagramm, das Sprachsignal-Wellenformen und die zugehörigen Veränderungen des dynamischen Meßwertes in der Ausgestaltung der Fig. 2 zeigt.

Fig. 5 ist ein Diagramm, das die Ergebnisse der Spracherfassung im Dokument nach Fig. 2 zeigt;

Fig. 6 ist ein Diagramm zum Erläutern einer anderen Ausgestaltung der vorliegenden Erfindung, die die Frequenz der spektralen Änderung mit einem Vektorquantisierungsschema kombiniert.

Fig. 7 ist ein Diagramm, das die Wirksamkeit der Ausgestaltung von Fig. 6 zeigt;

Fig. 8 ist ein Diagramm, das eine andere Ausgestaltung der vorliegenden Erfindung zeigt, bei der die Frequenz der spektralen Änderung mit der Tonhöhenfrequenz des Eingangssignals verknüpft sind; und

Fig. 9 ist ein Diagramm, das noch eine weitere Ausgestaltung der vorliegenden Erfindung zeigt, bei der die Frequenz der spektralen Änderung mit der Zahl von Nulldurchgängen des Eingangssignals verknüpft ist.

Gemäß der vorliegenden Erfindung wird eine spektrale Veränderung des Eingangssignals von einer Zeitfolge ihrer spektralen Merkmalparameter abgeleitet, und die zu erfassende Sprachperiode ist eine Periode, über der das Spektrum des Eingabesignals sich mit ungefähr der gleichen Frequenz wie die Sprachperiode ändert.

Die Erfassung einer Änderung im Spektrum des Eingangssignals beginnt mit dem Berechnen des Merkmalvektors des Spektrums zu jedem Zeitpunkt, gefolgt von einer Berechnung des dynamischen Merkmals aus dem Spektrum anhand von Merkmalvektoren an einer Mehrzahl von Punkten in der Zeit und dann durch Berechnen des Ausmaßes der Änderung im Spektrum aus der Norm des dynamischen Merkmalsvektors. Die Frequenz oder das zeitliche Muster der spektralen Veränderung im Sprachzeitraum ist vorberechnet, und eine Periode, in der das Eingangssignal eine spektrale Veränderung ähnlich der oben erwähnten erfährt, wird als Sprachperiode erfaßt.

Als spektraler Merkmalparameter kann Information über die spektrale Umhüllende benutzt werden, die durch eine FFT-Spektralanalyse, Cepstrum-Analyse, Kurzzeit-Autokorrelationsanalyse oder ähnliche Spektralanalyse erhältlich ist. Der spektrale Merkmalparameter ist üblicherweise eine Folge von mehreren Werten (entsprechend einer Folge von spektralen Frequenzen), die im folgenden als Merkmalsvektor bezeichnet wird. Das dynamische Merkmal kann die Differenz zwischen Zeitfolgen von spektralen Merkmalparametern, ein Polynom-Expansionskoeffizient oder beliebige andere spektrale Merkmalparameter sein, so lange sie die spektrale Veränderung darstellen. Die Frequenz der spektralen Veränderung wird durch ein Verfahren erfaßt, das in der Lage ist, den Grad der spektralen Änderung durch Zählen der Zahl von Peaks in der spektralen Veränderung über eine bestimmte Rahmenzeit oder durch Berechnen des Integrals des Ausmaßes der Änderung im Spektrum zu berechnen.

Natürlich ist ein Sprachgeräusch insbesondere eine Folge von Phonemen, und jedes Phonem hat eine charakteristische spektrale Umhüllende. Folglich ändert sich das Spektrum stark an der Grenze zwischen Phonemen. Außerdem ist die Zahl von Phonemen, die pro Zeiteinheit erzeugt werden (die Frequenz der Erzeugung der Phoneme) in einer solchen Folge von Phonemen nicht nach Sprachen unterschiedlich, sondern ist allgemeinen Sprachen gemeinsam. Bezogen auf die spektrale Veränderung kann das Sprachsignal charakterisiert werden als ein Signal, dessen Spektrum mit einer Periode nahezu gleich der Phonemlänge variiert. Diese Eigenschaft tritt in anderen Geräuschen in der natürlichen Welt nicht auf. Durch Vorausberechnen eines akzeptablen Bereichs der spektralen Veränderung in der Sprachperiode ist es möglich, als Sprachperiode eine Periode zu erfassen, in dem die Frequenz des Auftretens der spektralen Veränderung des Eingangssignals im vorberechneten Bereich liegt.

Als Verfahren zum Analysieren des Spektrums des Eingangssignals sind z.B. ein Verfahren zum direkten Frequenzanalysieren des Eingangssignals, ein FFT-(Fast Fourier-Transform)-Verfahren zum Analysieren des Eingangssignals und ein LPC-(Linear Predictive Coding)-Verfahren zum Analysieren des Eingangssignals bekannt. Es folgen Gleichungen zum Ableiten des spektralen Parameters nach drei repräsentativen Sprachspektralanalyseverfahren.

(a) Spektralparameter φ(m) durch Kurzzeit-Autokorrelationsanalyse:

$$\phi(m) = \frac{1}{N} \sum_{n=0}^{N-1-|m|} x(n)x(n+|m| \qquad (1)$$

(b) Spektralparameter S(ω) durch Kurzzeit-Spektralanalyse:

$$S(\omega) = \frac{1}{2\pi N} \left| \sum_{n=0}^{N-1} x(n)\exp(-j\omega n) \right|^2 \qquad (2)$$

(c) Spektralparameter Cn durch Cepstrum-Analyse:

$$Cn = \frac{1}{N}\sum_{k=0}^{N-1} \log|X(k)|\exp\{j2\pi kn/N\} \qquad (3)$$

Der Spektralparameter durch LPC-Cepstrum-Analyse wird in der gleichen Form wie Gleichung (3) ausgedrückt. Außerdem stellen ein linearer Vorhersagekoeffizient $\{\alpha_i | i=1, ..., p\}$, ein PARCOR-Koeffizient $\{K_i | i=1, ..., p\}$ und ein Linienspektrumpaar LSP ebenfalls Spektralhülleninformation von Sprachsignalen dar. Diese spektralen Parameter werden alle ausgedrückt durch eine Koeffizientenfolge (Vektor) und werden als akustische Merkmalvektoren bezeichnet. Eine Beschreibung wird typischerweise für das LPC-Cepstrum C = $\{c_1, c_2, ..., c_K\}$ angegeben, doch können auch andere spektrale Parameter verwendet werden.

Wie oben angegeben, ist das Prinzip der vorliegenden Erfindung, die Entscheidung, ob die Periode des Eingangssignals eine Sprachperiode ist, abhängig davon zu treffen, ob die Frequenz einer spektralen Änderung des Eingangssignals innerhalb eines vorgegebenen Bereiches liegt. Das Ausmaß der Änderung im Spektrum wird als dynamischer Messwert der Sprache wie unten beschrieben erhalten. Der erste Schritt ist, eine Zeitfolge von akustischen Parametervektoren des Sprachsignals durch FFT-Analyse, LPC-Analyse oder irgendeine andere Spektralanalyse zu erhalten. Nehmen wir an, daß ein k-dimensionales LPC-Cepstrum C(t) = $\{c_1, c_2, ..., c_k\}$ als Merkmalsvektor zum Zeitpunkt t verwendet wird. Um eine Änderung im Frequenzspektrum der Sprache über eine Fensterbreite 2n (wobei n die Zahl von diskreten Zeitpunkten ist) einer bestimmten Periode darzustellen, wird eine lokale Bewegung des Cepstrums C(t) durch ein gewichtetes Verfahren der kleinsten Fehlerquadrate linear approximiert, und ihre Neigung A(t) (ein linearer Differentialkoeffizient) wird als Ausmaß der Änderung im Spektrum (ein Gradientenvektor) erhalten. Das heißt, wenn die Gewichtung $w_i = w_{-i}$ gesetzt wird, ist die Neigung durch lineare Approximation gegeben durch die folgende Gleichung:

$$a_k(t) = \sum_{i=-n}^{n} iw_i c_k(t+i) \Big/ \sum_{i=-n}^{n} i^2 w_1 \qquad (4)$$

Dabei stellt $a_k(t)$ ein k-tes Element eines k-dimensionalen Vektors A(t) = $\{a_1(t), a_2(t), ..., a_k(t)\}$ dar, der das dynamische Merkmal des Spektrums zur Zeit t darstellt, und A(t) wird als ein Delta-Cepstrum bezeichnet. Das heißt, $a_k(t)$ bezeichnet einen linearen Differentialkoeffizienten einer Zeitfolge von k-dimensionalen Ceptstrumelementen $c_k(t)$ zur Zeit t (siehe Furui, "Digital Speech Processing", Tokai University Press).

Der dynamische Messwert D(t) zur Zeit t wird berechnet durch die folgende Gleichung, die die Summe der Quadrate aller Elemente des Delta-Cepstrums zur Zeit t darstellt (siehe Shigeki Sagayama and Fumitada Itakura, „On Individuality in a Dynamic Measure of Speech," Proc. Accoustical Society, Frühjahrskonferenz 1997, 3-3-7, Seiten 589 bis 590, Juni 1997).

$$D(t) = \sum_{k=1}^{K} a_k^2(t) \qquad (5)$$

Das heißt, das Cepstrum C(k) stellt das Merkmal der spektralen Hülle dar, und das Delata-C-Cepstrum, welches sein linearer Differentialkoeffizient ist, stellt das dynamische Merkmal dar. Der dynamische Messwert stellt also die Größe der spektralen Veränderung dar. Die Frequenz SF der spektralen Änderung wird berechnet als die Anzahl von Peaks der dynamischen Messwerte D(t), die im Laufe einer bestimmten Rahmenperiode F (eines Analyserahmens) einen vorgegebenen Schwellwert D_{th} überschreiten oder als Gesamtsumme (Integral) der dynamischen Messungen D(t) im Analyserahmen F.

Zwar ist oben der dynamische Messwert D(t) des Spektrums im Falle der Verwendung des Cepstrums C(t) als der spektrale Merkmals-(Vektor)-Parameter beschrieben worden, doch kann die dynamische Messung D(t) in ähnlicher Weise als andere spektrale Merkmalparameter definiert werden, die durch Vektoren dargestellt werden.

Sprache enthält z.B. zwei bis drei Phoneme in 400 Millisekunden, und das Spektrum variiert entsprechend der Zahl der Phoneme. Fig. 1 ist ein Graph, der die für viele Rahmen gemessene Zahl von Peaks zeigt, die starke Spektrumänderungen pro Zeiteinheit (400 ms, die als Analyserahmenlänge F definiert sind) anzeigen. 8 Stück Sprachdaten durch Lesen wurden verwendet. In Fig. 1 stellt die Abszisse die Zahl von Malen dar, wo die spektrale Veränderung einen Wert von 0,5 pro Rahmen überschritten hat, und die Ordinate stellt die Häufigkeit dar, mit der die jeweilige Zahl von Peaks gezählt wurde. Wie aus Fig. 1 offensichtlich ist, verteilt sich die Zahl von Peaks pro Rahmen zwischen 1 und 5. Diese Verteilung ändert sich zwar mit dem zum Bestimmen der Peaks verwendeten Schwellwert oder den verwendeten Sprachdaten, ist aber für Sprachgeräusche charakteristisch. Wenn das Spektrum des Eingangssignals in einer 400 ms-Periode ein- bis fünfmal variiert, kann somit entschieden werden, daß eine Sprachsignalperiode vorliegt. Die Änderung im Spektrum (Merkmalsvektor) stellt die Neigung der Zeitfolge C(t) der Merkmalvektoren an jedem Zeitpunkt dar.

Fig. 2 zeigt eine Ausgestaltung der vorliegenden Erfindung. Ein über einen Signaleingabeanschluß 11 eingegebenes Signal S wird in einem A/D-Wandlerteil 12 in ein digitales Signal gewandelt. Ein Extraktionsteil für akustisches Merkmal 13 berechnet das akustische Merkmal des gewandelten digitalen Signals wie etwa dessen LPC- oder FFT-Cepstren. Ein Berechnungsteil für einen dynamischen Messwert 14 berechnet das Ausmaß der Änderung im Spektrum aus der LPC-Cepstrenfolge. Das heißt, das LPC-Cepstrum wird alle 10 ms erhalten, indem die LPC-Analyse des Eingangssignals für jedes Analysefenster von z.B. 20 ms Breite durchgeführt wird, wie in Zeile A in Fig. 3 gezeigt, wodurch eine Folge von LPC-Cepstren C(0), C(1), C(2), ..., erhalten wird, wie in Zeile B in Fig. 3 gezeigt. Jedesmal wenn das LPC-Cepstrum C(t) erhalten wird, wird das Delta-Cepstrum A(t) nach Gleichung (4) aus den $2n+1$ letzten LPC-Cepstren berechnet, wie in Zeile C in Fig. 3 gezeigt. Fig. 3 zeigt den Fall, wo n gleich 1 ist. Als nächstes wird jedesmal, wenn das Delta-Cepstrum A(t) erhalten wird, das dynamische Maß D(t) nach Gleichung (5) berechnet, wie in Zeile D in Fig. 3 gezeigt.

Indem die oben beschriebene Verarbeitung über den Analyserahmen F von 400 ms Zeitlänge durchgeführt wird, von dem angenommen wird, daß er eine Mehrzahl von Phonemen umfaßt, werden 40 dynamische Messungen D(t) erhalten. Ein Sprachperioden-Erfassungsteil 15 zählt die Zahl von Peaks der dynamischen Messwerte D(t), die den Schwellwert D_{th} überschreiten und liefert den Zählwert als Frequenz S_F der Spektrumsänderung.

Alternativ wird die Gesamtsumme der dynamischen Messwerte D(t) über den Analyserahmen F berechnet und als Frequenz S_F der Spektrumänderung definiert.

Die Frequenz der Spektrumänderung in der Sprachperiode wird vorausberechnet, auf deren Grundlage der obere und untere Schwellwert vorgegeben werden. Der Rahmen des Eingangssignals, der in den Bereich zwischen dem unteren und dem oberen Schwellwert fällt, wird als ein Sprachrahmen erfaßt. Schließlich wird das Sprachperioden-Erfassungsergebnis aus einem Sprachperioden-Erfassungsausgabeteil ausgegeben. Indem die Frequenz S_F der Spektrumsänderung während der Anwendung des Eingangssignals wiederholt durchgeführt und dabei die zeitliche Position des Analyserahmens F jedesmal um ein Zeitintervall von 20 ms verschoben wird, wird die Sprachperiode im Eingangssignal erfaßt.

Fig. 4 ist ein Diagramm, das eine Sprachsignal-Wellenform und ein Beispiel eines Musters der entsprechenden Änderung der dynamischen Messung D(t) zeigt. Die in Zeile A gezeigten Sprachwellenformdaten sind die Aussprache, durch einen männlichen Sprecher, der japanischen Wörter /keikai/ und /sasuga/, mit der Bedeutung "Achtung" bzw. "wie zu erwarten". Die LPC-Cepstrumanalyse zum Erhalten des dynamischen Messwerts D(t) des Eingangssignals wurde durchgeführt mit einem 20 ms langen Analysefenster, das um ein 10 ms-Zeitintervall verschoben wurde. Das Delta-Cepstrum A(t) wurde über einer Rahmenbreite von 100 ms berechnet. Aus Fig. 4 ist zu sehen, daß der dynamische Messwert D(t) in einem stillen Bereich oder stationären Bereich der Sprache nicht stark variiert, wie in Zeile B gezeigt, und daß Peaks der dynamischen Messungen an Anfangs- und Endpunkten der Sprache oder an der Grenze zwischen Phonemen auftreten.

Fig. 5 ist ein Diagramm zur Erläuterung eines Beispiels des Ergebnisses der Erfassung von Sprache mit überlagertem Geräusch. Die in Zeile A gezeigte Eingangssignal-Wellenform wurde wie folgt erzeugt: das Geräusch eines fahrenden Autos wurde mit einem Signal-Rausch-Verhältnis von 0 dB einem Signal überlagert, das durch Verkettung der Aussprache des japanischen Wortes /aikawarazu/ mit der Bedeutung "wie üblich" durch zwei Sprecher erhalten wurde, wobei die Aussprachen jeweils durch eine stille Periode von 5 s getrennt waren. Zeile B in Fig. 5 zeigt eine korrekte Sprachperiode, die die Periode darstellt, in der Sprache vorhanden ist. Zeile D zeigt Änderungen in der dynamischen Messung D(t). Zeile C zeigt das automatisch auf der Basis von Änderungen des dynamischen Messwerts D(t) automatisch ermittelte Sprachperioden-Erfassungsergebnis. Der dynamische Messwert D(t) wurde unter den gleichen Bedingungen wie in Fig. 4 erhalten. Folglich wurde der dynamische Messwert alle 10 ms erhalten. Die Analyserahmenlänge war 400 ms, und der Analyserahmen wurde in Schritten von 200 ms verschoben. Die Gesamtsumme der dynamischen Messwerte D(t) in der Analyserahmenperiode wurde als Frequenz S_F der Spektrumänderung berechnet. In diesem Beispiel wurde der Analyserahmen F,

für den der Wert dieser Gesamtsumme einen vorgegebenen Wert von 4,0 überschritt, als Sprachperiode erfaßt. Während Sprachperioden auf der Eingangssignal-Wellenform wegen des niedrigen Signal-Rausch-Verhältnisses nicht klar zu sehen sind, ist zu sehen, daß mit dem erfindungsgemäßen Verfahren alle Sprachperioden erfaßt wurden. Fig. 5 zeigt, daß die vorliegende Erfindung die Frequenz der Spektrumänderung ausnutzt und so die Erfassung von Sprache im Rauschen ermöglicht.

Fig. 6 ist ein Diagramm zur Erläuterung einer anderen Ausgestaltung der vorliegenden Erfindung, die sowohl den dynamischen Messwert als auch die Spektralhülleninformation nutzt, um die Sprachperiode zu erfassen. Wie bei der oben erwähnten Ausgestaltung der Fall ist, wird das über den Signaleingangsanschluß 11 eingegebene Signal vom A/D-Wandlerteil 13 in ein digitales Signal umgesetzt. Das Extraktionsteil 13 berechnet für das gewandelte digitale Signal das akustische Merkmal wie etwa das LPC- oder FFT-Cepstrum. Das Rechenteil 14 für den dynamischen Messwert berechnet den dynamischen Messwert D(t) auf der Grundlage des akustischen Merkmals. Ein Vektorquantisierer 17 nimmt Bezug auf einen Vektorquantisierungs-Codebuchspeicher 18, liest dann daraus vorberechnete repräsentative Vektoren von Sprachmerkmalen aus und berechnet Vektorquantisierungsverzerrungen zwischen den repräsentativen Vektoren und Merkmalvektoren des Eingangssignals, um so die minimale Quantisierungsverzerrung zu erfassen. Wenn das Eingangssignal im Analysefenster ein Sprachsignal ist, kann der zu diesem Zeitpunkt erhaltene akustische Merkmalsvektor ein mit einem relativ kleinen Ausmaß an Verzerrung quantisierter Vektor sein, indem auf das Codebuch des Vektorquantisierungs-Codebuchspeichers 18 zurückgegriffen wird. Wenn jedoch das Eingangssignal im Analysefenster kein Sprachsignal ist, erzeugt die Vektorquantisierung ein großes Ausmaß an Verzerrung. So ist es durch Vergleichen der Vektorquantisierungsverzerrung mit einem vorgegebenen Pegel von Verzerrung möglich, zu entscheiden, ob das Eingangssignal in dem Sprachanalysefenster ein Sprachsignal oder nicht ist.

Das Sprachperioden-Erfassungsteil 15 entscheidet, daß ein Signal über die 400 ms-Analyserahmenperiode ein Sprachsignal ist, wenn die Frequenz S_F der Änderung des dynamischen Messwerts in den durch den oberen und unteren Grenzwert definierten Bereich fallen und die Quantisierungsverzerrung zwischen dem Merkmalvektor und dem Eingangssignal und dem entsprechenden repräsentativen Sprachmerkmalvektor kleiner als ein vorgegebener Wert ist. Diese Ausgestaltung verwendet zwar die Vektorquantisierungsverzerrung, um das Merkmal der spektralen Hülle zu untersuchen, es ist jedoch auch möglich, eine zeitliche Folge von vektorquantisierten Codes zu verwenden, um zu bestimmen, ob eine für Sprache charakteristische Sequenz darunter ist. Außerdem kann auch manchmal ein Verfahren zum Erhalten eines Sprach-Entscheidungsraumes in einem spektralen Merkmalraum verwendet werden.

Es folgt eine Beschreibung eines Beispiels eines Experimentes, in dem Sprache durch eine Kombination des dynamischen Maßes und des Sprachmerkmalvektors erfaßt wird, die die oben erwähnte Vektorquantisierungsverzerrung minimiert. Dies ist ein Beispiel für ein Experiment zum Erfassen von Sprache aus einem Eingangssignal, das aus Sprache und dem Singen eines Vogels im Wechsel miteinander zusammengesetzt ist. Im Experiment wurde das Vektorquantisierungscodebuch aus einer großen Menge von Sprachdaten erzeugt. Als Sprachdaten wurden die

Aussprachen von 50 Worten und 25 Sätzen durch 20 Sprecher aus einer ATR-Sprachdatenbank ausgewählt. Die Zahl von Quantisierungspunkten ist 512. Der Merkmalvektor ist ein 16-dimensionales LPC-Cepstrum, die Analysefensterbreite ist 30 ms, und die Fensterverschiebungsbreite ist 10 ms. Die Summe von Quantisierungsverzerrungen von alle 10 ms gelieferten Merkmalvektoren wurde berechnet unter Verwendung des in Schritten von 200 ms verschobenen, 400 ms langen Analysefensters. Entsprechend wurde die Summe der dynamischen Messwerte ebenfalls unter Verwendung des in Schritten von 200 ms verschobenen, 400 ms langen Analysefensters berechnet. Für den dynamischen Messwert wie auch für die Quantisierungsverzerrung ist der Bereich ihrer akzeptablen Werte in der Sprachperiode basierend auf dem Lernen von Sprache voreingestellt, und die Sprachperiode wird erfaßt, wenn eingegebene Sprache in den Bereich fällt.

Das zur Bewertung verwendete Eingangssignal waren abwechselnde Verkettungen von 8 Sätzen, jeweils aufgebaut aus ca. 5 Sekunden langer Sprache, und 8 Arten von Vogelgesang von jeweils 5 Sekunden Länge, ausgewählt aus einer Datenbank für kontinuierliche Sprache der Japanischen Akustischen Gesellschaft. Die folgenden Maße werden gesetzt, um die Leistung dieser Ausgestaltung zu bewerten.

Rahmenerfassungsrate = (Anzahl von korrekt erfaßten Sprachrahmen)/(Anzahl von Sprachrahmen in den Bewertungsdaten)

Richtig-Rate = (Anzahl von korrekt erfaßten Sprachrahmen)/(Anzahl von vom System als Sprache ausgegebenen Rahmen)

Die Richtig-Rate stellt das Ausmaß dar, in dem das vom System als Sprachrahmen angegebene Ergebnis korrekt ist. Die Erfassungsrate stellt das Ausmaß dar, in dem das System Sprachrahmen im Eingangssignal erfassen konnte. In Fig. 7 sind unter Verwendung der obigen Messwerte die Ergebnisse der Spracherfassung mit Bezug auf die Bewertungsdaten gezeigt. Die Änderungsgeschwindigkeit des Spektrums des Vogelgesanges hat eine starke Ähnlichkeit mit der Änderungsgeschwindigkeit des Spektrums der Sprache; deshalb wird, wenn nur der dynamische Messwert verwendet wird, Vogelgesang so oft irrtümlich als Sprache erfaßt, daß die Richtig-Rate niedrig ist. Durch die kombinierte Verwendung des dynamischen Messwerts und der Vektorquantisierungsverzerrung kann die spektrale Hülle des Vogelgesanges von der spektralen Hülle von Sprache unterschieden werden, und die Richtig-Rate nimmt entsprechend zu.

Im Falle eines langen Vokals wie etwa eines Diphthongs kann das Spektrum manchmal in der Vokalperiode keine Veränderungen erfahren. Wenn Sprache einen solchen Vokal enthält, besteht eine Möglichkeit eines Erfassungsfehlers, die nur mit dem erfindungsgemäßen Verfahren auftritt, bei dem die Spektrumsänderung genutzt wird. Indem dieses erfindungsgemäße Verfahren mit der bislang verwendeten Erfassung der Tonhöhenfrequenz, des Amplitudenwertes oder des Autokorrelationskoeffizienten des Eingangssignals kombiniert wird, ist es möglich, die Möglichkeit zu verringern, daß dieser Erfassungsfehler auftritt. Die Tonhöhenfrequenz ist die Zahl von Schwingungen der menschlichen Stimmbänder und reicht von 50 bis 500 Hz und tritt im stationären Teil des Vokals deutlich auf. Das heißt, die Tonhöhenfrequenzkomponente hat üblicherweise eine starke Amplitude (Leistung), und das Vorhandensein der Tonhöhenfrequenzkomponente bedeutet, daß der Wert des Autokorrelationskoeffizienten in dieser Periode groß ist. Durch Erfassen der Anfangs- und Endpunkte und der Periodizität der Sprachperiode über die Erfassung der Frequenz

der Spektrumänderung nach diesem erfindungsgemäßen Verfahren und durch Erfassen des Vokalteils mit der Tonhöhenfrequenz und/oder der Amplitude und/oder dem Autokorrelationskoeffizienten ist es möglich, die Möglichkeit von Erfassungsfehlern zu reduzieren, die im Falle von einen langen Vokal enthaltender Sprache auftreten.

Fig. 8 zeigt eine andere Ausgestaltung der vorliegenden Erfindung, die die Ausgestaltung der Fig. 2 mit dem Vokalerfassungsschema kombiniert. Die Schritte 12 bis 16 in Fig. 8 werden nicht beschrieben, da sie jenen in Fig. 2 entsprechen. Ein Vokalerfassungsteil 21 erfaßt z.B. die Tonhöhenfrequenz. Der Vokalerfassungsteil 21 erfaßt die Tonhöhenfrequenz im Eingangssignal und liefert sie an das Sprachperiodenerfassungsteil 15. Das Sprachperiodenerfassungsteil 15 bestimmt in der gleichen Weise wie oben, ob die Frequenz S_F der Änderung des dynamischen Messwerts D(t) im vorgegebenen Schwellwertbereich ist, und entscheidet, ob die Tonhöhenfrequenz in dem für menschliche Sprache typischen Bereich von 50 bis 500 Hz liegt. Ein Eingangssignalrahmen, der diese zwei Bedingungen erfüllt, wird als ein Sprachrahmen erfaßt. In Fig. 8 ist gezeigt, daß das Vokalerfassungsteil 21 getrennt von den Hauptverarbeitungsschritten 12 bis 16 vorgesehen ist, da aber in der Praxis die Tonhöhenfrequenz, die spektrale Leistung oder der Autokorrelationswert durch Berechnung in Schritt 13 im Rahmen der Cepstrumberechnung erhalten werden können, muß der Vokalerfassungsteil 21 nicht immer getrennt vorgesehen sein. Während in Fig. 8 gezeigt ist, daß die Erfassung der Tonhöhenfrequenz für die Erfassung der Sprachvokalperiode genutzt ist, ist es auch möglich, die Tonhöhenfrequenz und/oder die Leistung und/oder den Autokorrelationswert zu berechnen und sie für die Entscheidung über das Sprachsignal zu nutzen.

Für die Erfassung der Sprachperiode kann die in Fig. 8 gezeigte Vokalerfassung durch die Erfassung eines Konsonanten ersetzt werden. Fig. 9 zeigt eine Kombination der Erfassung der Anzahl von Nulldurchgängen und der Erfassung der Frequenz der Spektrumsänderung. Stimmlose Reiblaute haben meist eine Verteilung von 400 bis 1.400 Nulldurchgängen pro Sekunde. Folglich ist es möglich, ein Verfahren zu verwenden, das den Anfangspunkt eines Konsonanten erfaßt, indem ein geeigneter, von einem Nulldurchgangsanzahl-Erfassungsteil 22 ausgewählter Schwellwert der Nulldurchgangsanzahl verwendet wird, wie in Fig. 9 gezeigt.

Das erfindungsgemäße, oben beschriebene Sprachperioden-Erfassungsverfahren kann angewendet werden auf einen Sprachschalter, der ein Gerät sprachgesteuert ein- oder ausschaltet, oder auf die Erfassung von Sprachperioden für die Spracherkennung. Außerdem ist das erfindungsgemäße Verfahren anwendbar auf das Auffinden von Sprache in Videoinformation oder akustischen CD-Informationsdaten.

Da erfindungsgemäß wie oben beschrieben die Sprachperiode auf der Grundlage der Frequenz der der für menschliche Sprache charakteristischen Spektrumsänderung erfaßt wird, kann die Sprachperiode sogar aus Sprache stabil erfaßt werden, der Rauschen mit hoher Leistung überlagert ist. Auch kann ein Geräusch mit einem der Sprache ähnlichen Leistungsmuster als Nicht-Sprache erkannt werden, wenn die Geschwindigkeit seiner Spektrumsänderung sich von der Phonemschaltgeschwindigkeit der Sprache unterscheidet. Deshalb ist die vorliegende Erfindung anwendbar auf die Erfassung der Sprachperiode, die bei der Vorverarbeitung wiederer-

kannt werden muß, wenn eine Spracherkennungseinheit in stark verrauschter Umgebung verwendet wird, oder z.B. auf die Technik zum Wiederfinden einer Konversationsszene aus akustischen Daten eines Fernsehprogramms, Spielfilms oder ähnlichen Medien, die Musik oder diverse Geräusche enthalten sowie auf das Editieren eines Videos und Zusammenfassen von dessen Inhalt. Außerdem ermöglicht die vorliegende Erfindung die Erfassung der Sprachperiode mit höherer Genauigkeit durch Kombinieren der Frequenz der Spektrumsänderung mit dem Leistungswert, der Nulldurchgangsanzahl, dem Autokorrelationskoeffizienten oder der Grundfrequenz, die ein anderes Merkmal von Sprache ist.

Es liegt auf der Hand, daß diverse Abwandlungen und Änderungen durchgeführt werden können, ohne den Rahmen der neuartigen Konzepte der vorliegenden Erfindung, wie in den nachfolgenden Ansprüchen definiert, zu verlassen.

PATENTANSPRÜCHE

1. Signalverarbeitungsverfahren zum Erfassen einer Sprachperiode in einem Eingangssignal, mit den Schritten:

(a) Erhalten eines spektralen Merkmalparameters durch Analysieren des Spektrums des Eingangssignals für jedes vorgegebene Analysefenster;

(b) Berechnen des Ausmaßes der Änderung des spektralen Merkmalparameters des Eingangssignals pro Zeiteinheit;

(c) Berechnen der Änderungsfrequenz des Ausmaßes der Änderung des spektralen Merkmalparameters über eine vorgegebene Analyserahmenperiode, die länger als die Zeiteinheit ist; und

(d) Überprüfen, ob die Frequenz der Änderung in einen vorgegebenen Frequenzbereich fällt, und, wenn ja, Entscheiden, daß das Eingangssignal des Analyserahmens ein Sprachsignal ist.

2. Verfahren nach Anspruch 1, bei dem der Schritt des Berechnens des Ausmaßes der Änderung des spektralen Merkmalparameters einen Schritt des Erhaltens einer Zeitfolge von Merkmalvektoren, die die Spektren des Eingangssignals an jeweiligen Zeitpunkten darstellen, und einen Schritt des Berechnens von dynamischen Merkmalen durch Verwendung der Merkmalvektoren an einer Mehrzahl von Zeitpunkten und des Berechnens der Änderung im Spektrum des Eingangssignals aus der Norm der dynamischen Merkmale umfaßt.

3. Verfahren nach Anspruch 2, bei dem das dynamische Merkmal Polynom-Expansionskoeffizienten der Merkmalvektoren an einer Mehrzahl von Zeitpunkten sind.

4. Verfahren nach Anspruch 1, 2 oder 3, bei dem der Schritt des Berechnens der Frequenz ein Schritt des Zählens der Anzahl der einen vorgegebenen Schwellwert überschreitenden Peaks der Spektrumsänderung in dem Analyserahmen und des Lieferns des Zählwertes als die Frequenz ist.

5. Verfahren nach Anspruch 1, 2 oder 3, bei dem der Schritt des Berechnens der Frequenz einen Schritt des Berechnens der Gesamtsumme der Änderungen im Spektrum des Eingangssignals in der vorgegebenen Analyserahmenperiode, die länger als die Zeiteinheit ist, umfaßt, und der Schritt des Entscheidens entscheidet, daß das Eingangssignal der Analyserahmenperiode ein Sprachsignal ist, wenn die Gesamtsumme in einen vorgegebenen Wertebereich fällt.

6. Verfahren nach Anspruch 4 oder 5, soweit nicht auf Anspruch 3 bezogen, bei dem der Schritt des Berechnens der Spektrumsänderung einen Schritt des Berechnens eines Gradientenvektors, als dessen Elemente lineare Differentialkoeffizienten von jeweiligen Elementen eines den spektralen Merkmalparameter darstellenden Vektors verwendet werden, und einen Schritt des Berechnens von Quadratsummen der jeweiligen Elemente des Gradientenvektors als dynamische Messwerte der Spektrumsänderung umfaßt.

7. Verfahren nach Anspruch 6, bei dem der spektrale Merkmalparameter ein LPC-Cepstrum ist und die Spektrumänderung ein Delta-Cepstrum ist.

8. Verfahren nach Anspruch 1, ferner mit einem Schritt des vektoriellen Quantisierens des Eingangssignals für jedes der Analysefenster durch Bezugnahme auf ein Vektorcodebuch, das aufgebaut ist aus aus Sprachdaten erhaltenen repräsentativen Vektoren von spektralen Merkmalparametem von Sprache, und des Berechnens von Quantisierungsverzerrung, wobei in dem Schritt des Entscheidens entschieden wird, daß das Eingangssignal ein Sprachsignal ist, wenn die Quantisierungsverzerrung kleiner als ein vorgegebener Wert ist und die Frequenz der Änderung innerhalb des vorgegebenen Frequenzbereiches liegt.

9. Verfahren nach Anspruch 1, ferner mit einem Schritt des Erfassens, ob das Eingangssignal in einem jeweiligen Analysefenster ein Vokal ist, und wobei in dem Entscheidungsschritt (d) entschieden wird, ob das Eingangssignal ein Sprachsignal ist, indem ein Vokal erfaßt wird und erfaßt wird, ob die Frequenz der Änderung in dem vorgegebenen Frequenzbereich liegt.

10. Verfahren nach Anspruch 9, bei dem in dem Vokalerfassungsschritt eine Tonhöhenfrequenz in dem Eingangssignal für jedes Analysefenster erfaßt wird und entschieden wird, daß das Eingangssignal ein Vokal ist, wenn die erfaßte Tonhöhenfrequenz in einem vorgegebenen Frequenzbereich liegt.

11. Verfahren nach Anspruch 9, bei dem in dem Vokalerfassungsschritt die Leistung des Eingangssignals für jedes Analysefenster erfaßt wird und entschieden wird, daß das Eingangssignal ein Vokal ist, wenn die erfaßte Leistung größer als ein vorgegebener Wert ist.

12. Verfahren nach Anspruch 9, bei dem in dem Vokalerfassungsschritt der Autokorrelationswert des Eingangssignals erfaßt wird und entschieden wird, daß das Eingangssignal ein Vokal ist, wenn der erfaßte Autokorrelationswert größer als ein vorgegebener Wert ist.

13. Verfahren nach Anspruch 1, ferner mit einem Schritt (e) des Zählens der Anzahl von Nulldurchgängen des Eingangssignals in jedem Analysefenster und des Entscheidens, daß das Eingangssignal in dem Analysefenster ein Konsonant ist, wenn der Zählwert innerhalb eines vorgegebenen Bereiches liegt, und wobei in dem Entscheidungsschritt (d) entschieden wird, ob das Eingangssignal Sprache ist, indem durch den Entscheidungsschritt (e) entschieden wird, ob

das Eingangssignal ein Konsonant ist und entschieden wird, ob die Änderungsfrequenz in dem vorgegebenen Frequenzbereich liegt.

14. Verfahren nach Anspruch 1, 2 oder 3, bei dem der spektrale Merkmalparameter ein LPC-Cepstrum ist.

15. Verfahren nach Anspruch 1, 2 oder 3, bei dem der spektrale Merkmalparameter ein FFT-Cepstrum ist.

EP 0 764 937

FIG.1

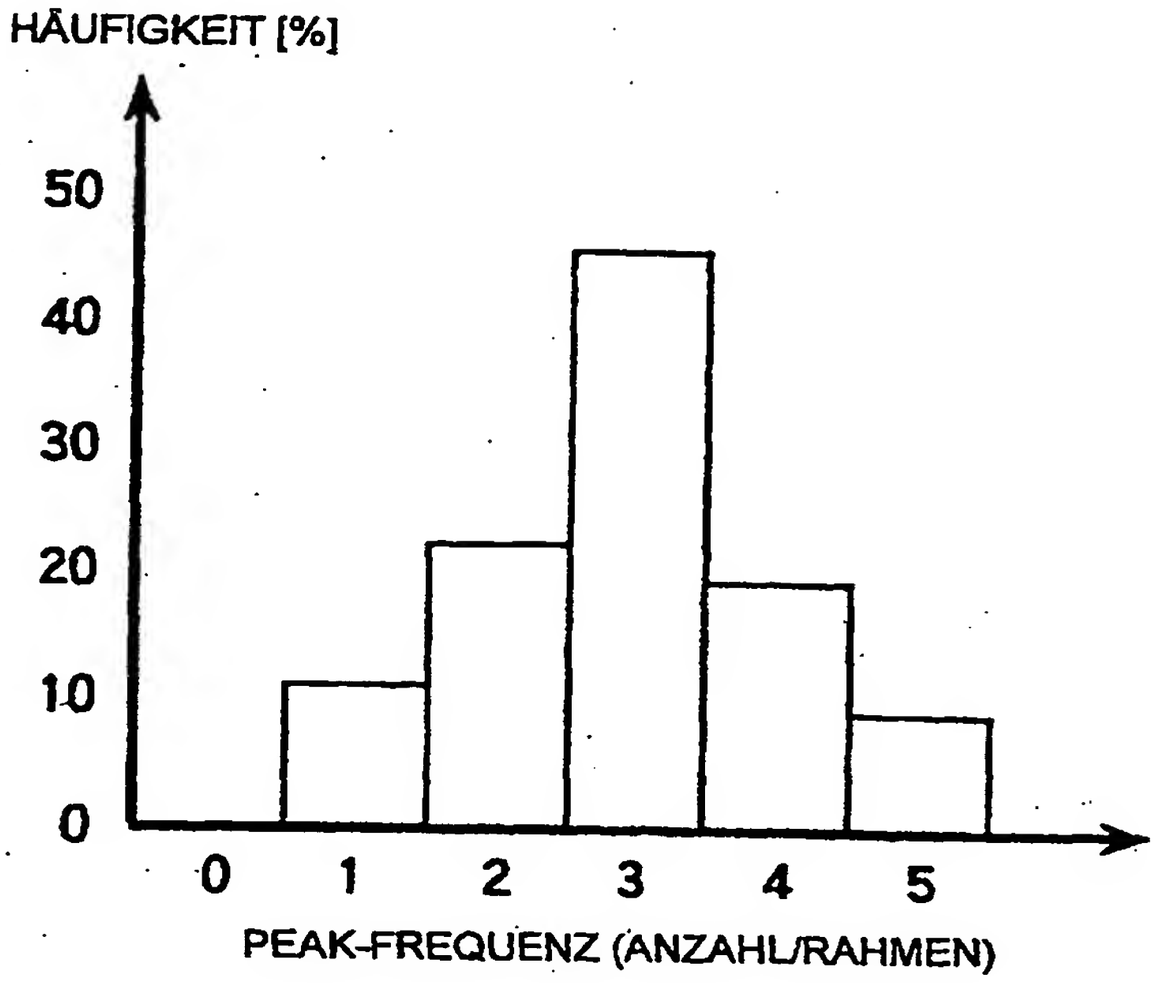


FIG.2

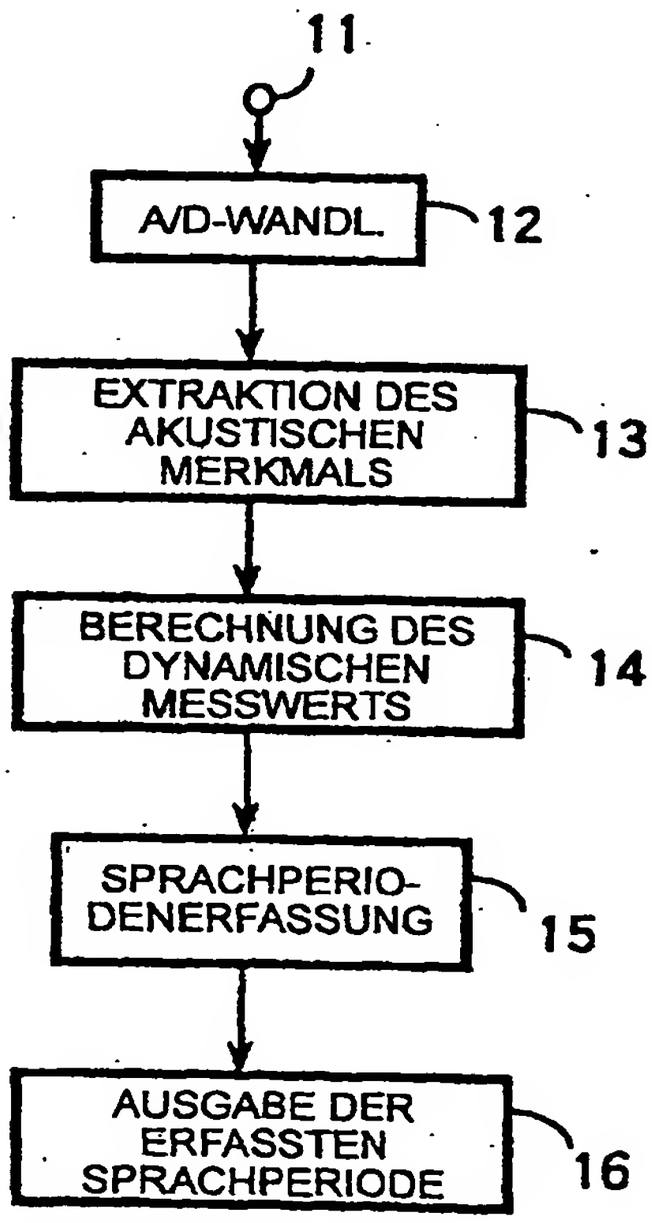


FIG. 3

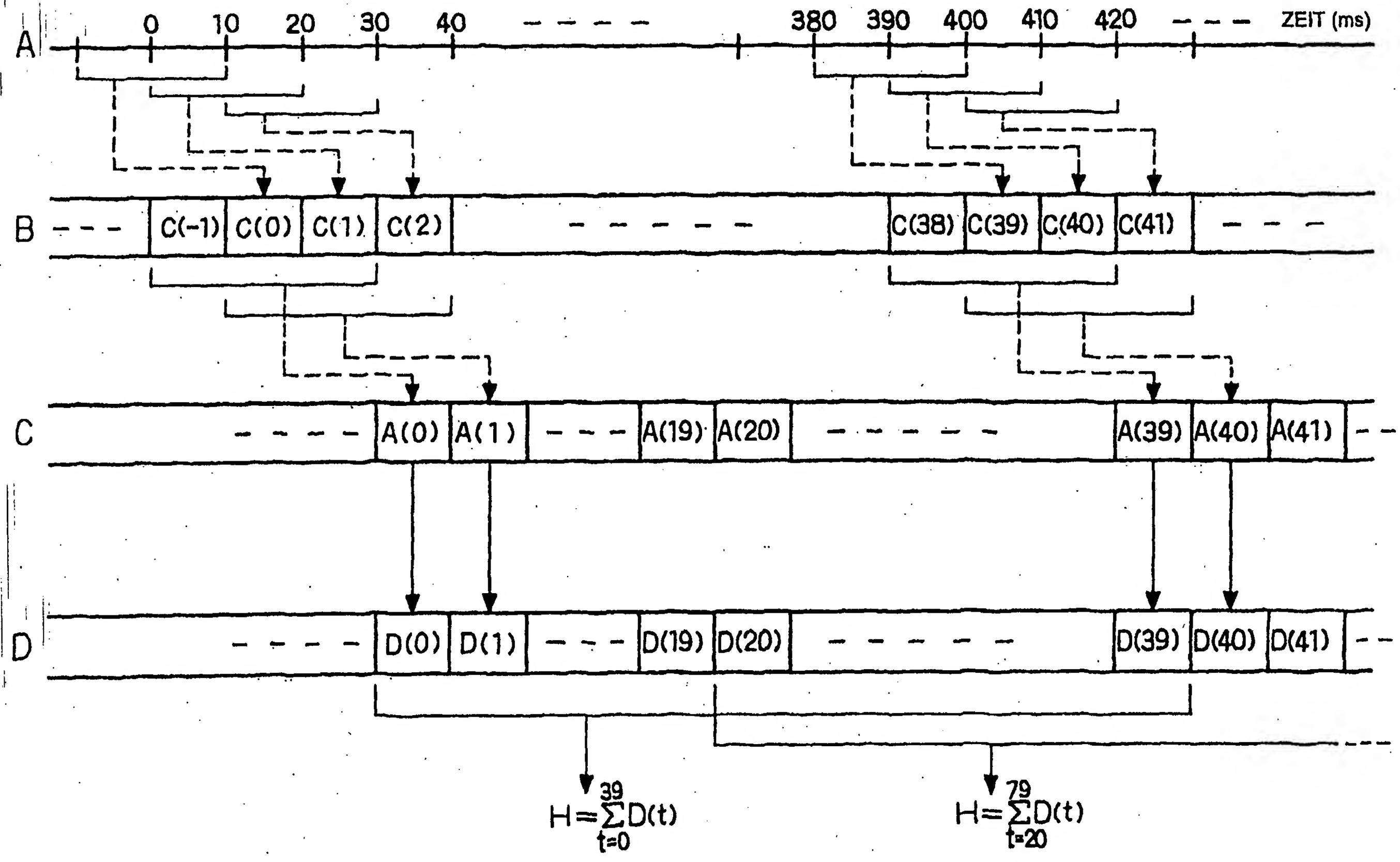

A
0
10
20
30
40
380
390
400
410
420
ZEIT (ms)
B
C(-1)
C(0)
C(1)
C(2)
C(38)
C(39)
C(40)
C(41)
C
A(0)
A(1)
A(19)
A(20)
A(39)
A(40)
A(41)
D
D(0)
D(1)
D(19)
D(20)
D(39)
D(40)
D(41)
H=Σ D(t) t=0 39
H=Σ D(t) t=20 79

FIG. 4

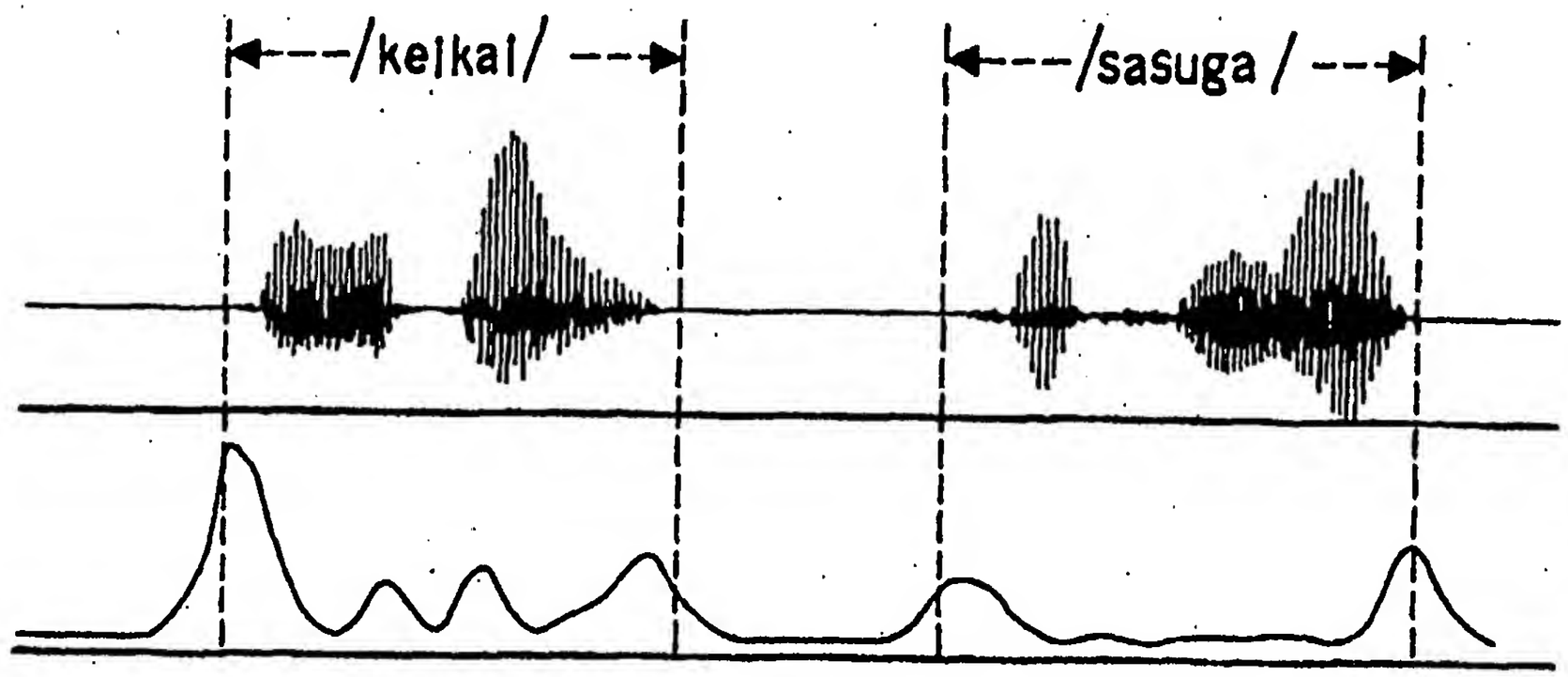

A
SPRACHSIGNAL-
WELLENFORM
B
DYNAMISCHER
MESSWERT
/kɛlkaɪ/
/sasuga /

FIG. 5

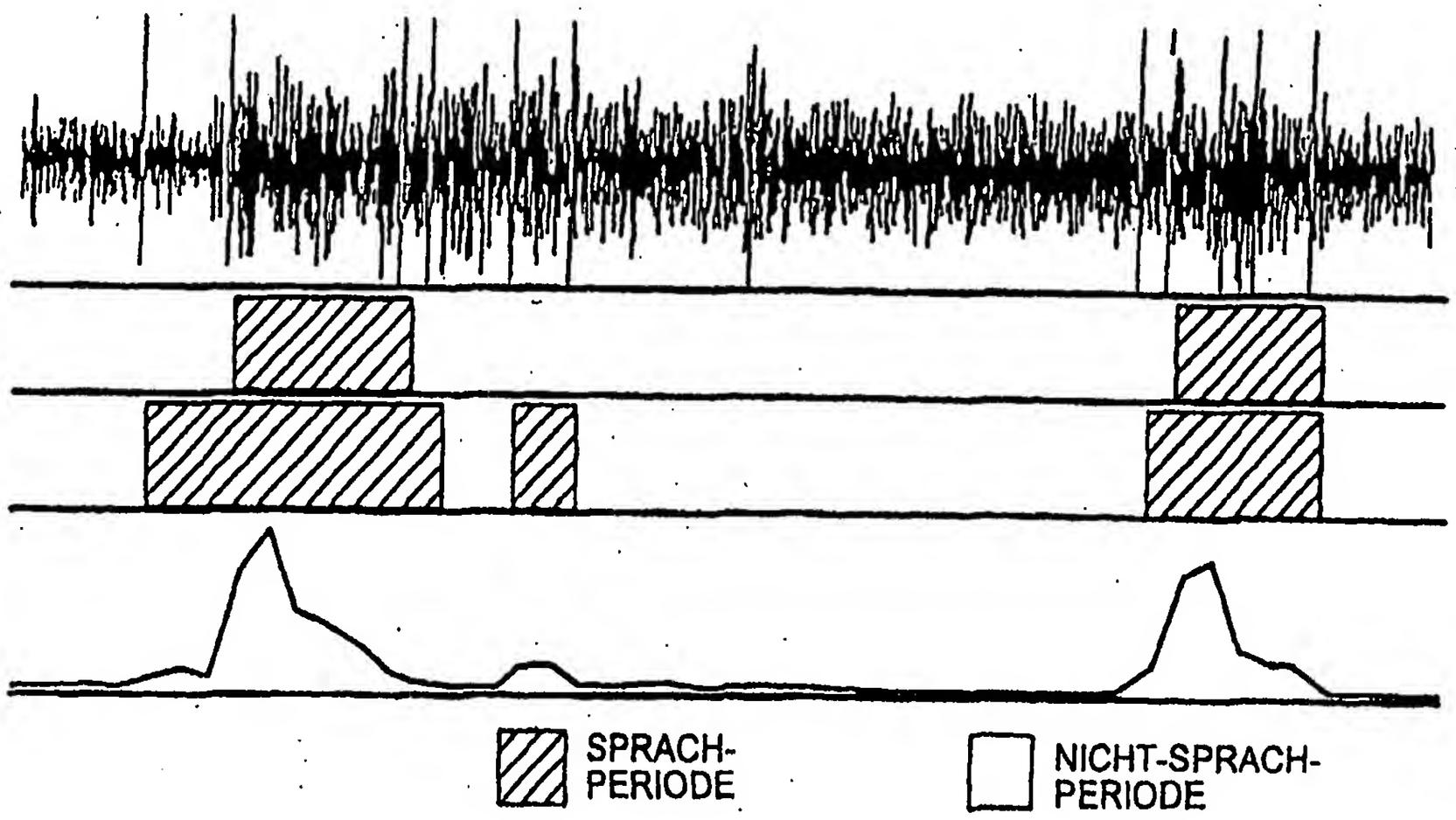

A
EINGANGSSIGNAL-
WELLENFORM
B
KORREKTE
SPRACHPERIODE
C
ERFASSUNGSPERIODE
D
DYNAMIKÄNDERUNG
SPRACH-
PERIODE
NICHT-SPRACH-
PERIODE

FIG.6

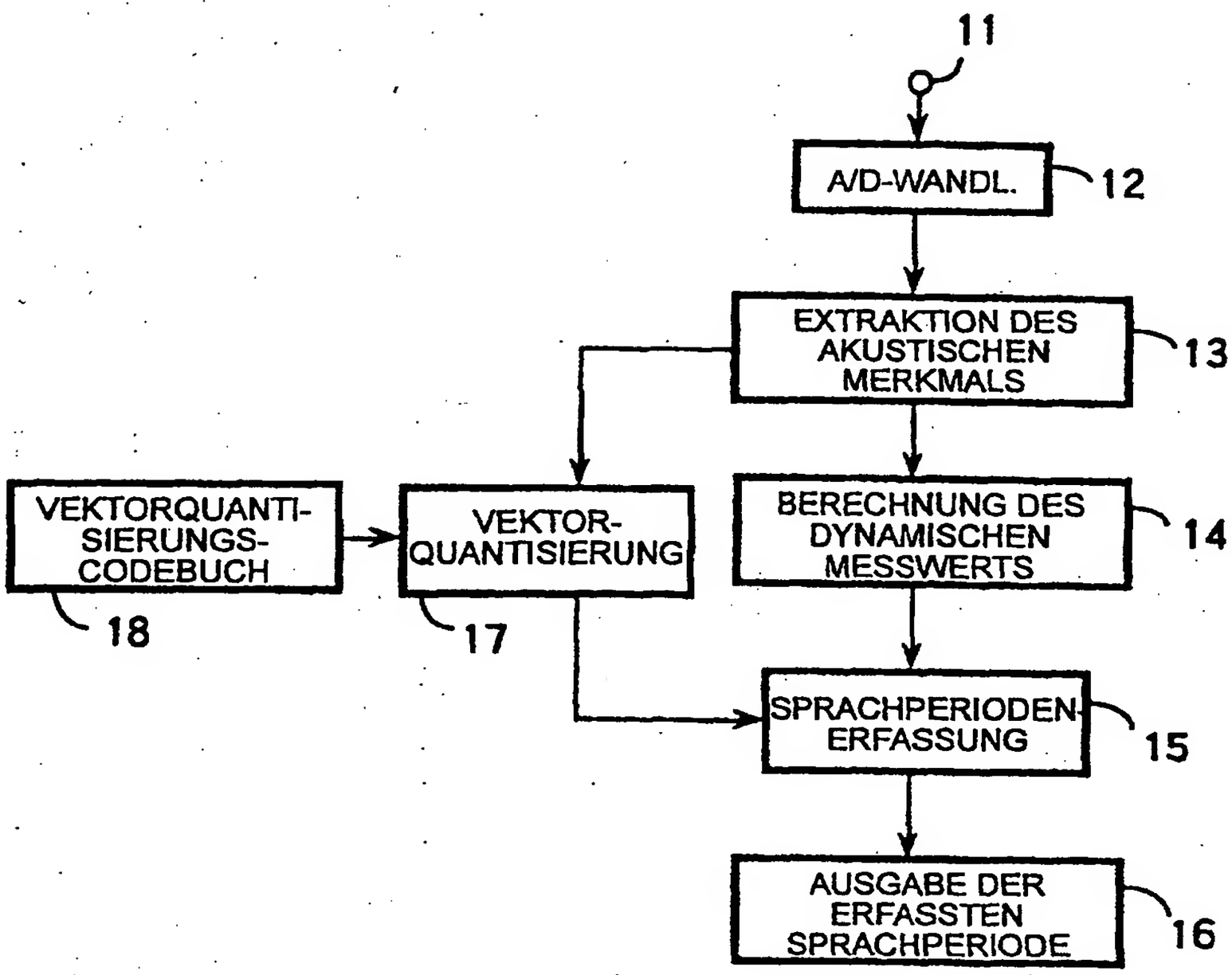


FIG.7

	ERFASSUNGSRATE [%]	RICHTIG-RATE [%]
NUR DYNAMISCHER MESSWERT	84.4	34.6
DYNAMISCHER MESSWERT UND QUANTISIERUNGS-VERZERRUNG	83.3	80.0

FIG.8

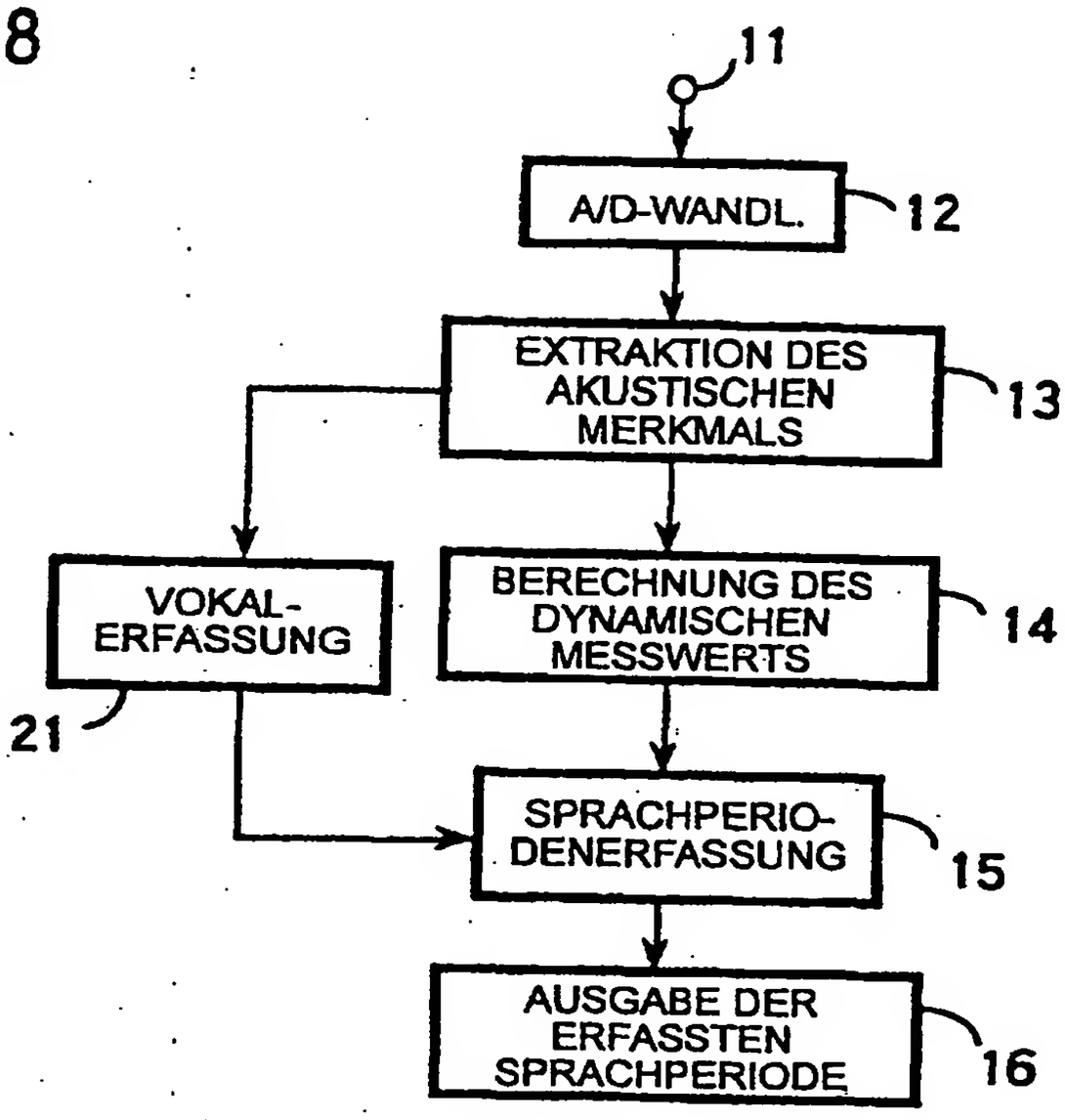
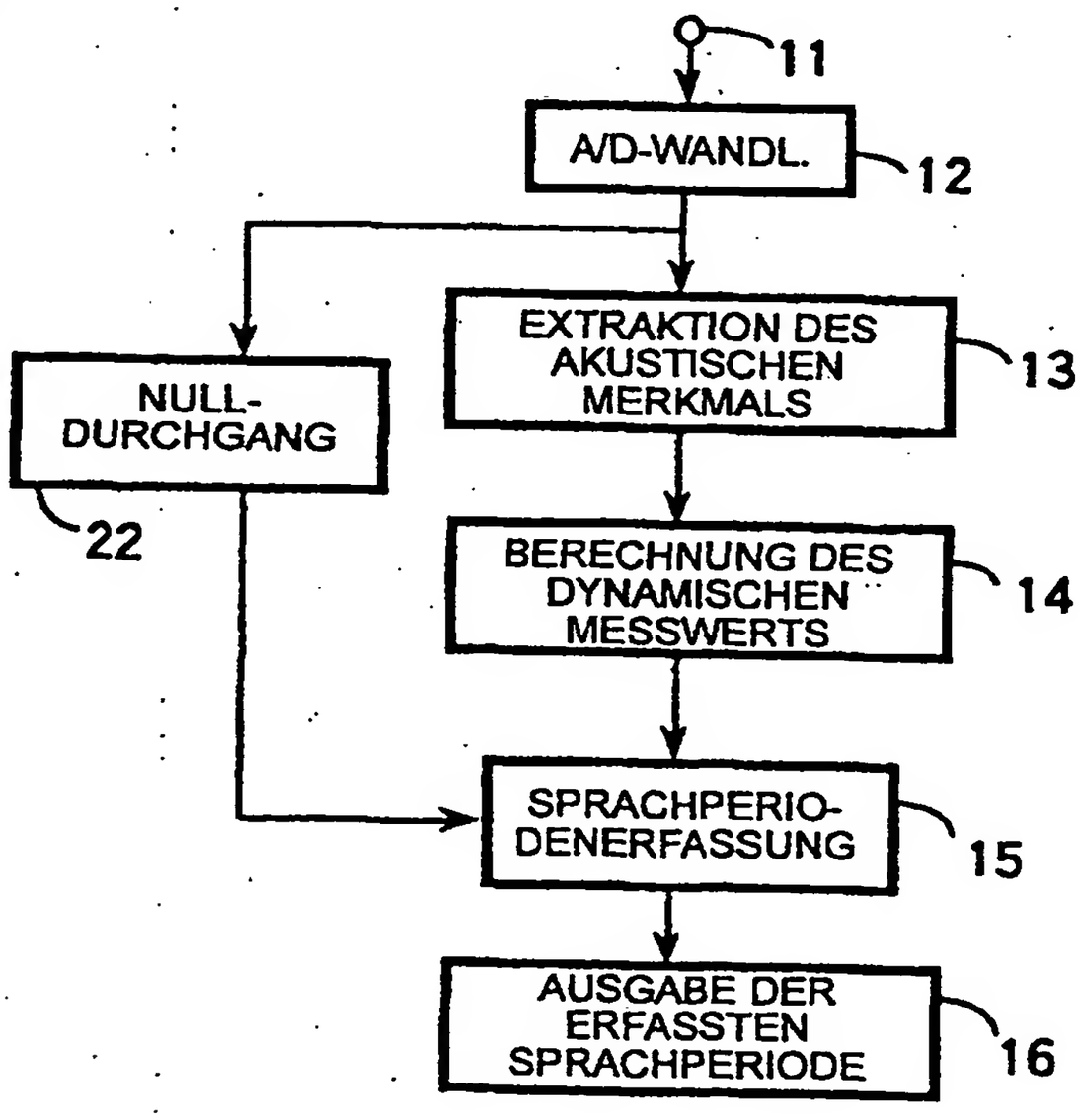

11
A/D-WANDL.
12
EXTRAKTION DES AKUSTISCHEN MERKMALS
13
VOKAL-ERFASSUNG
21
BERECHNUNG DES DYNAMISCHEN MESSWERTS
14
SPRACHPERIO-DENERFASSUNG
15
AUSGABE DER ERFASSTEN SPRACHPERIODE
16

FIG.9


11
A/D-WANDL.
12
NULL-DURCHGANG
22
EXTRAKTION DES AKUSTISCHEN MERKMALS
13
BERECHNUNG DES DYNAMISCHEN MESSWERTS
14
SPRACHPERIO-DENERFASSUNG
15
AUSGABE DER ERFASSTEN SPRACHPERIODE
16

Docket # 2004P00324

Applic. #

Applicant: T. Fingscheidt,

Lerner Greenberg Stemer LLP et al.
Post Office Box 2480
Hollywood, FL 33022-2480
Tel: (954) 925-1100 Fax: (954) 925-1101

Method for speech detection in a high-noise environment

Publication number: DE69613646T

Publication date: 2002-05-16

Inventor: MIZUNO OSAMU (JP); TAKAHASHI SATOSHI (JP); SAGAYAMA SHIGEKI (JP)

Applicant: NIPPON TELEGRAPH & TELEPHONE (JP)

Classification:

- international: ***G10L11/02;*** **G10L11/00;** (IPC1-7): G10L11/02; G10L15/20

- european: G10L11/02

Application number: DE19966013646T 19960923

Priority number(s): JP19950246418 19950925

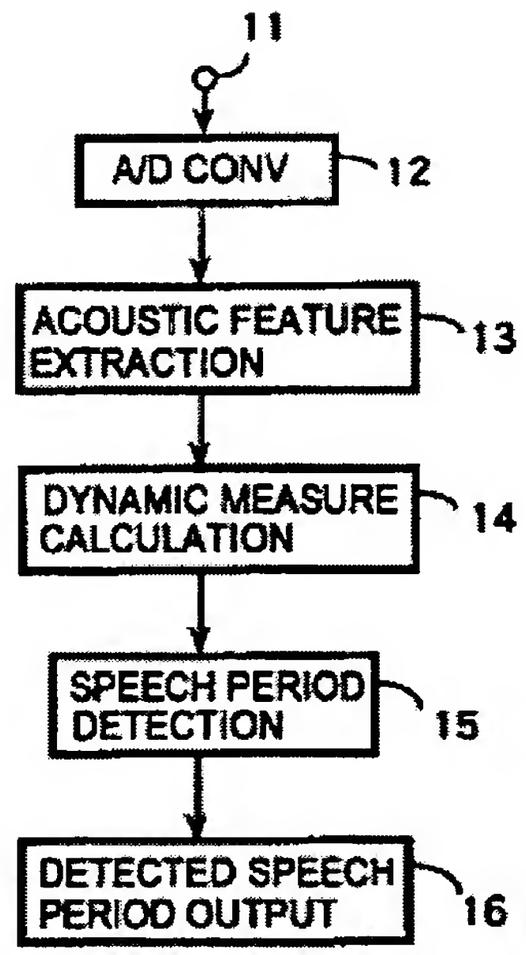
Also published as:

EP0764937 (A:
US5732392 (A
JP9090974 (A)
EP0764937 (A:
EP0764937 (B

Abstract not available for DE69613646T
Abstract of corresponding document: **EP0764937**

In method for detecting a speech period in a high-noise environment, the variation in the spectrum of an input signal per unit time is calculated over an analysis frame period, and when the frequency of spectrum variation falls in a predetermined range, the input signal of that frame is decided to be a speech signal.

FIG.2



Docket # 2004P00324
Applic. #
Applicant: T. Fingscheidt, et al.
Lerner Greenberg Stemer LLP
Post Office Box 2480
Hollywood, FL 33022-2480
Tel: (954) 925-1100 Fax: (954) 925-1101